

Supplemental Material

Argonaute CLIP-Seq reveals miRNA targetome diversity across tissue types

**Peter M. Clark, Phillip Loher, Kevin Quann,
Jonathan Brody, Eric R. Londin and Isidore Rigoutsos**

Sample ID	PMID or GEO accession	Species	Tissue	Percent Uniquely Mapped Reads
MMU Brain rep_01 (110 kDa)	19536157	Mus Musculus	P13 mouse neocortex	50.2
MMU Brain rep_02 (110 kDa)	19536157	Mus Musculus	P13 mouse neocortex	50.9
MMU Brain rep_03 (110 kDa)	19536157	Mus Musculus	P13 mouse neocortex	51.0
MMU Brain rep_01 (130 kDa)	19536157	Mus Musculus	P13 mouse neocortex	70.8
MMU Brain rep_02 (130 kDa)	19536157	Mus Musculus	P13 mouse neocortex	69.7
MMU Brain rep_03 (130 kDa)	19536157	Mus Musculus	P13 mouse neocortex	66.9
MMU mESC rep_01	21258322	Mus Musculus	Embryonic Stem Cells	54.2
MMU mESC rep_02	21258322	Mus Musculus	Embryonic Stem Cells	45.7
MMU mESC rep_03	21258322	Mus Musculus	Embryonic Stem Cells	36.2
MMU CD4 ⁺ T-cells rep_01	23142080	Mus Musculus	CD4 ⁺ T-cells	36.5
MMU CD4 ⁺ T-cells rep_02	23142080	Mus Musculus	CD4 ⁺ T-cells	18.1
MMU CD4 ⁺ T-cells rep_03	23142080	Mus Musculus	CD4 ⁺ T-cells	25.4
MMU CD4 ⁺ T-cells rep_04	23142080	Mus Musculus	CD4 ⁺ T-cells	23.4
MMU CD4 ⁺ T-cells rep_05	23142080	Mus Musculus	CD4 ⁺ T-cells	26.3
MMU CD4 ⁺ T-cells rep_06	23142080	Mus Musculus	CD4 ⁺ T-cells	24.3
MMU CD4 ⁺ T-cells rep_07	23142080	Mus Musculus	CD4 ⁺ T-cells	6.3
MMU CD4 ⁺ T-cells rep_08	23142080	Mus Musculus	CD4 ⁺ T-cells	25.9
MMU CD4 ⁺ T-cells rep_09	23142080	Mus Musculus	CD4 ⁺ T-cells	36.1
MMU CD4 ⁺ T-cells rep_10	23142080	Mus Musculus	CD4 ⁺ T-cells	15.3
MMU CD4 ⁺ T-cells rep_11	23142080	Mus Musculus	CD4 ⁺ T-cells	14.2
MMU CD4 ⁺ T-cells rep_12	23142080	Mus Musculus	CD4 ⁺ T-cells	14.1
MMU CD4 ⁺ 155KO T-cells rep_01	23142080	Mus Musculus	CD4 ⁺ T-cells	41.7
MMU CD4 ⁺ 155KO T-cells rep_02	23142080	Mus Musculus	CD4 ⁺ T-cells	32.0
MMU CD4 ⁺ 155KO T-cells rep_03	23142080	Mus Musculus	CD4 ⁺ T-cells	25.7
MMU CD4 ⁺ 155KO T-cells rep_04	23142080	Mus Musculus	CD4 ⁺ T-cells	31.2
MMU CD4 ⁺ 155KO T-cells rep_05	23142080	Mus Musculus	CD4 ⁺ T-cells	23.9
MMU CD4 ⁺ 155KO T-cells rep_06	23142080	Mus Musculus	CD4 ⁺ T-cells	26.8
MMU CD4 ⁺ 155KO T-cells rep_07	23142080	Mus Musculus	CD4 ⁺ T-cells	16.5
MMU CD4 ⁺ 155KO T-cells rep_08	23142080	Mus Musculus	CD4 ⁺ T-cells	17.2
MMU CD4 ⁺ 155KO T-cells rep_09	23142080	Mus Musculus	CD4 ⁺ T-cells	31.1
MMU CD4 ⁺ 155KO T-cells rep_10	23142080	Mus Musculus	CD4 ⁺ T-cells	28.7
MMU CD4 ⁺ 155KO T-cells rep_11	23142080	Mus Musculus	CD4 ⁺ T-cells	13.8
MMU CD4 ⁺ 155KO T-cells rep_12	23142080	Mus Musculus	CD4 ⁺ T-cells	16.8
HSA HEK293 rep_01	21572407	Homo Sapiens	Embryonic Kidney	20.5
HSA HEK293 rep_02	21572407	Homo Sapiens	Embryonic Kidney	6.5
HSA hTERT-HPNE	SRP034075	Homo Sapiens	hTERT-HPNE	18.7
HSA MIA PaCa-2	SRP034075	Homo Sapiens	MIA PaCa-2	12.4

Supp. Table 1. Comprehensive table of samples analysed within this study along with their Pubmed ID or GEO accession, species, tissue type and the percentage of uniquely mapped reads.

	Total # of HITS-CLIP sites (FDR <= 0.05)	Total # of Par-Clip sites (FDR <= 0.05)	Overlap
Ago CLIP complete T1 repA	7040	6883	97.77%
Ago CLIP complete T1 repB	13	11	84.62%

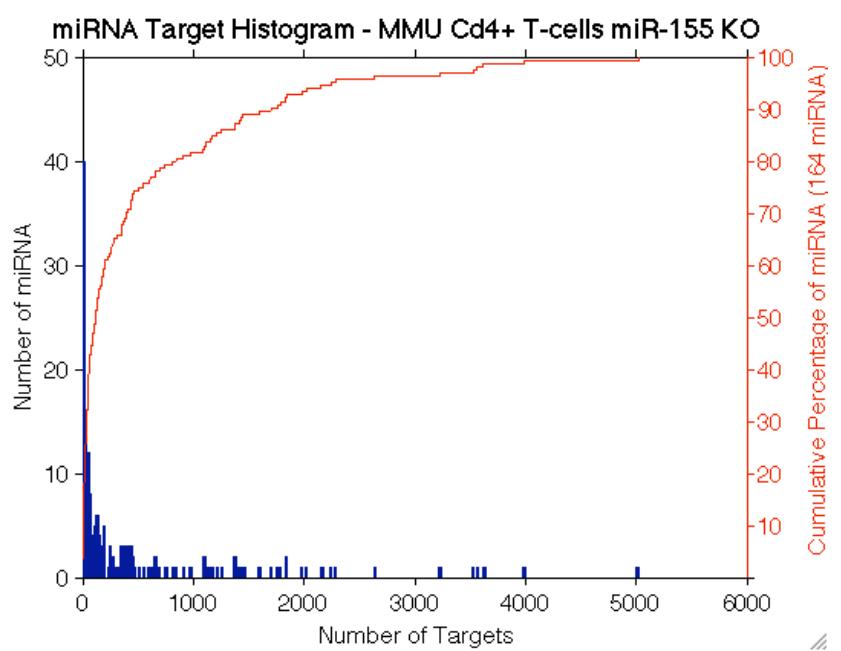
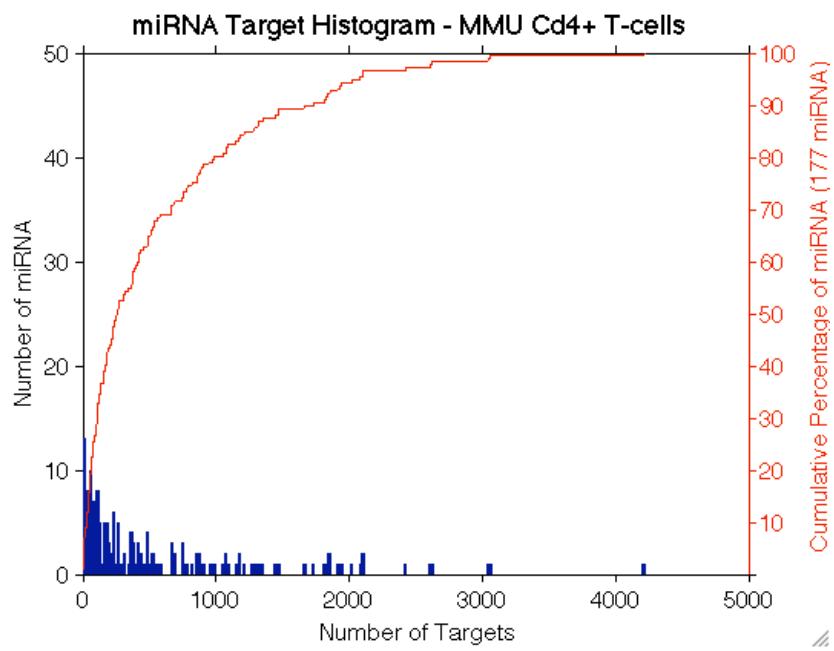
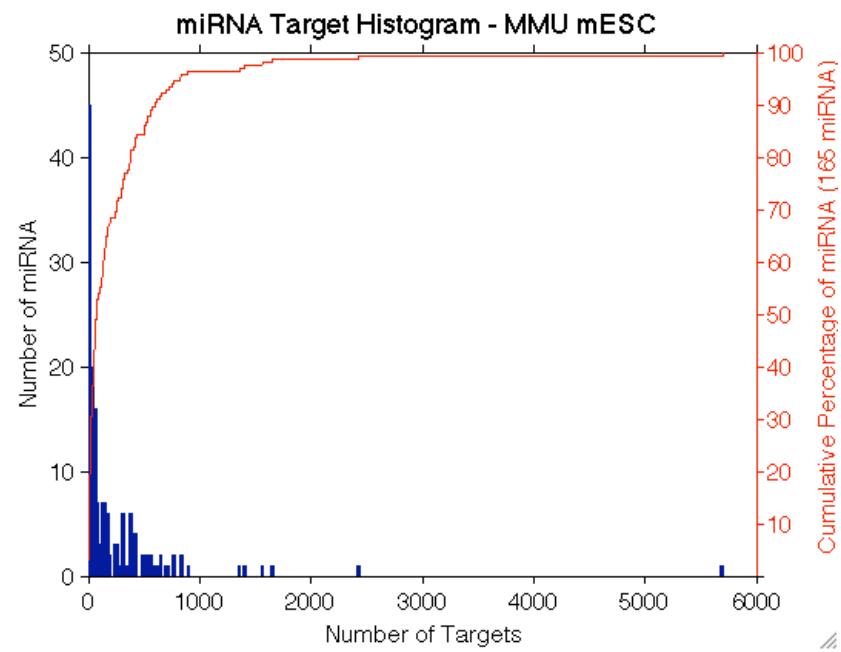
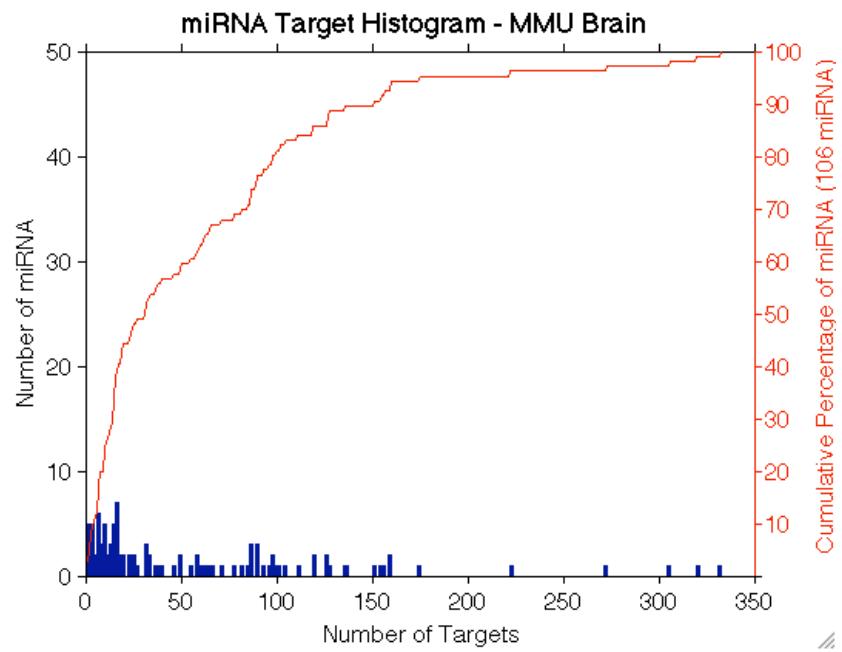
Supp. Table 2. Agreement of statistically significant MRE sites that are deduced by processing a HITS-CLIP and a PAR-CLIP dataset from the same cell type (HEK293). Original data from [1].

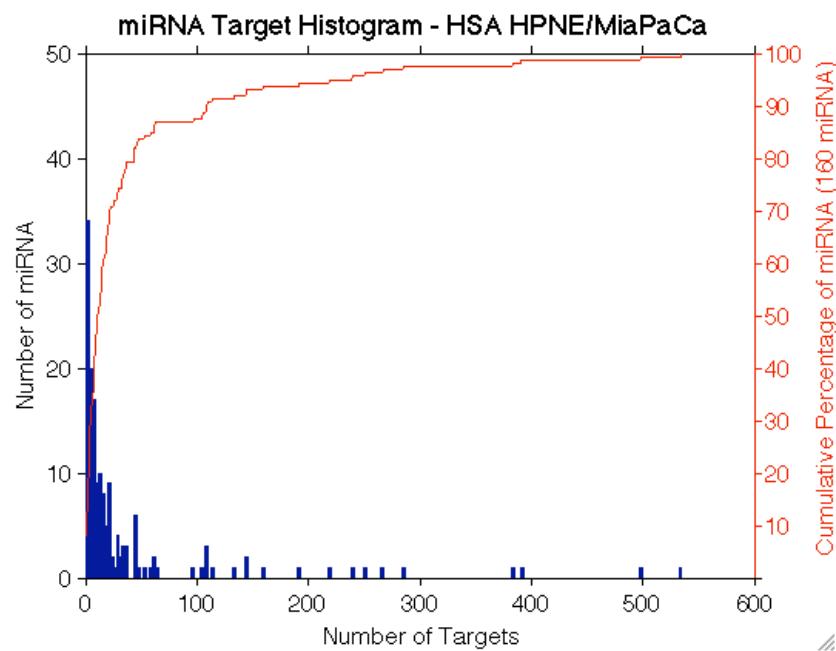
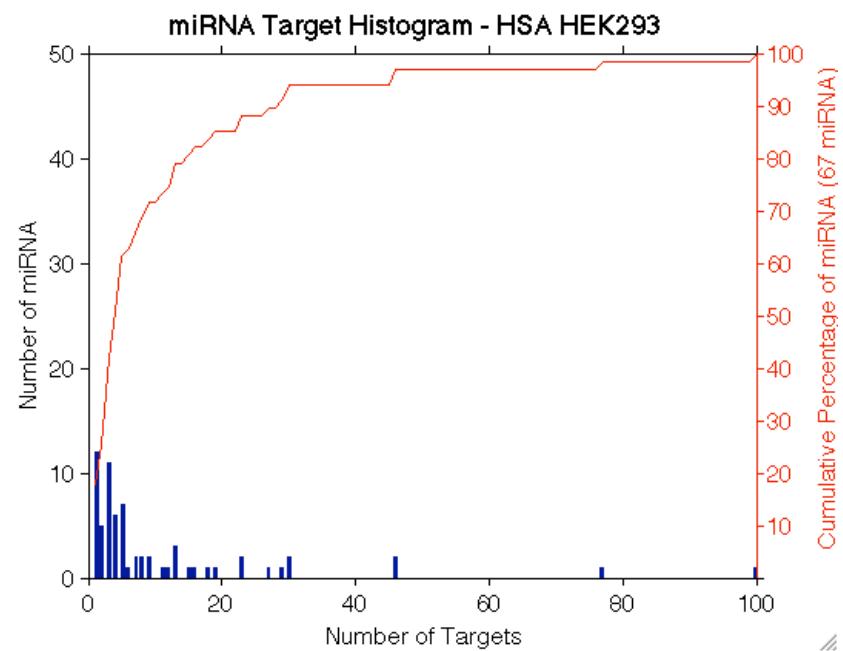
GO Biological Process	Fold Enrichment	FDR
regulation of transcription	1.7	2.8E-08
transcription	1.8	8.9E-08
proteolysis involved in cellular protein catabolic process	2.5	3.3E-06
cellular protein catabolic process	2.5	3.9E-06
protein catabolic process	2.5	4.5E-06
cellular macromolecule catabolic process	2.4	4.9E-06
modification-dependent protein catabolic process	2.6	5.2E-06
modification-dependent macromolecule catabolic process	2.6	5.2E-06
macromolecule catabolic process	2.3	2.3E-05
chromatin organization	2.9	1.5E-04
hemopoietic or lymphoid organ development	2.9	1.3E-03
intracellular signaling cascade	1.9	1.4E-03
chromosome organization	2.4	2.2E-03
immune system development	2.7	3.5E-03
regulation of programmed cell death	2.1	4.4E-03
chromatin modification	2.9	4.7E-03
regulation of cell death	2.1	5.1E-03
phosphorus metabolic process	1.8	8.9E-03
phosphate metabolic process	1.8	8.9E-03
phosphorylation	1.9	1.3E-02
hemopoiesis	2.8	1.4E-02
regulation of apoptosis	2.1	1.8E-02
negative regulation of cell proliferation	2.9	2.0E-02
positive regulation of macromolecule metabolic process	2.0	2.1E-02
regulation of transcription from RNA polymerase II promoter	2.0	2.3E-02
regulation of leukocyte activation	3.3	3.1E-02
regulation of cell activation	3.3	3.7E-02
lymphocyte differentiation	3.7	3.8E-02
leukocyte differentiation	3.4	3.9E-02
cellular macromolecule localization	2.5	4.2E-02
leukocyte activation	2.8	4.3E-02

Supp. Table 3. Enriched GO biological processes for miR-155 targets identified by our analysis.

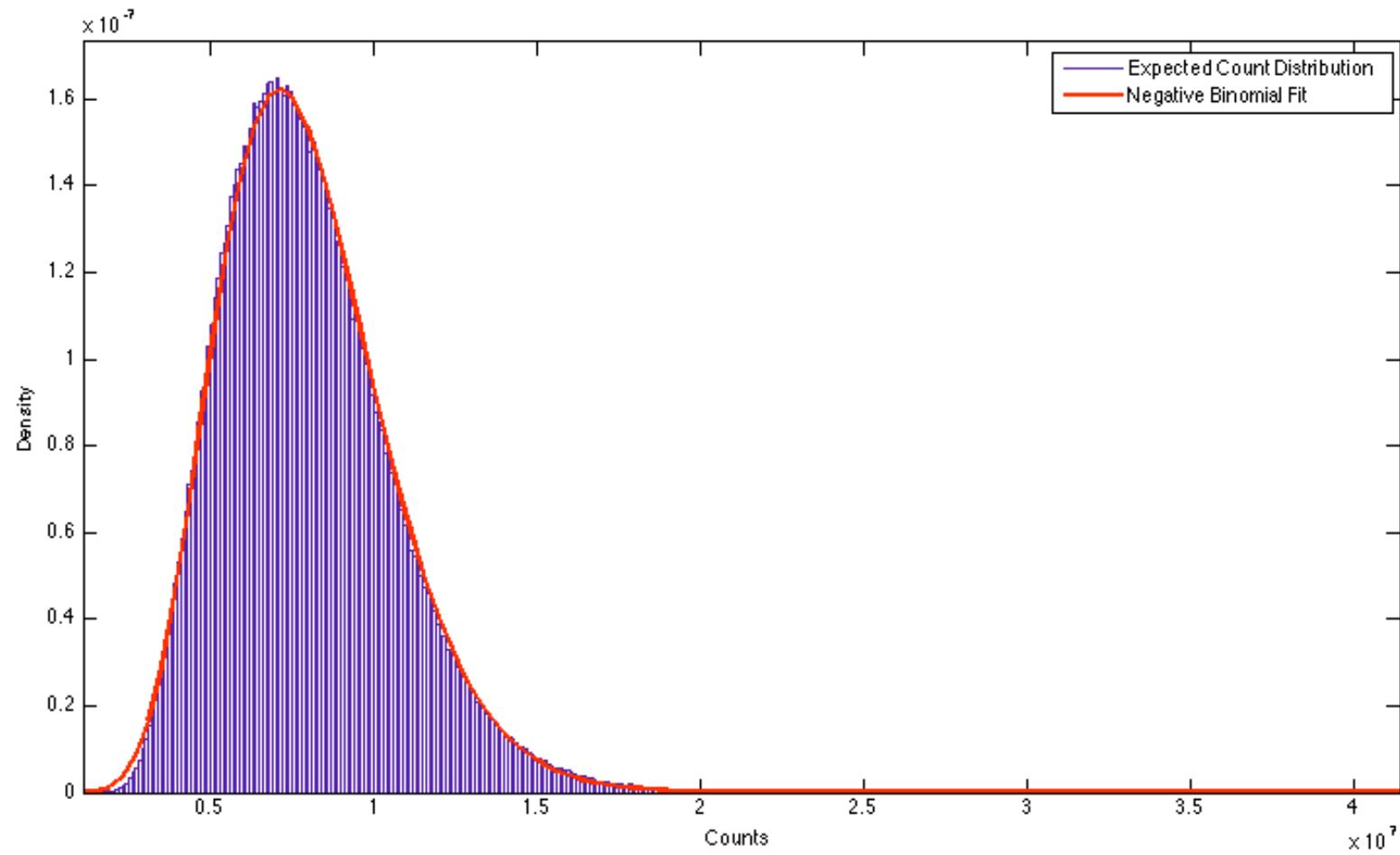
Species	# starbase targets	# starbase targets located within 3'UTR	# standard-model predictions within 3'UTR using those miRs that are present in Starbase	# overlapping miR/MRE prediction within 3'UTR	% overlap with Starbase
Human	601,189	557,128	3,079	2,355	76.49%
Mouse	111,809	103,955	28,471	12,047	42.31%

Supp. Table 4. Percentage of overlap between Starbase and our predictions (FDR < 0.05). Only canonical targets located in 3'UTR are considered.



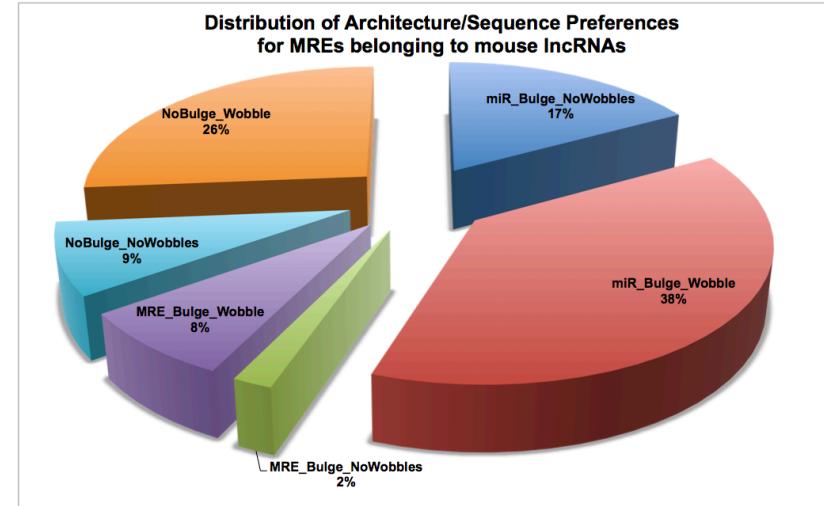
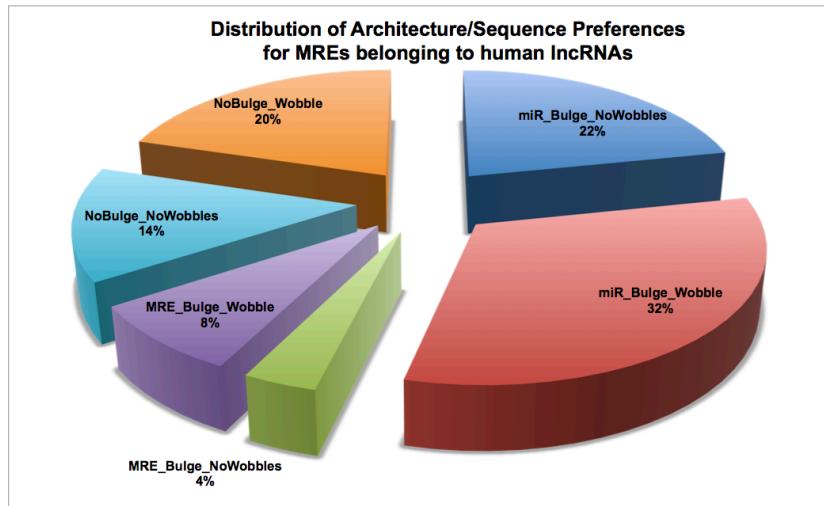


Supp. Figure 1. Distribution of the number of targets for the miRNAs participating in heteroduplexes where the MREs are targeted by a single miRNA. The miRNA:MRE heteroduplexes from replicates of the same samples have been pooled together (union of miRNA:MRE interactions across replicates from a given sample type).



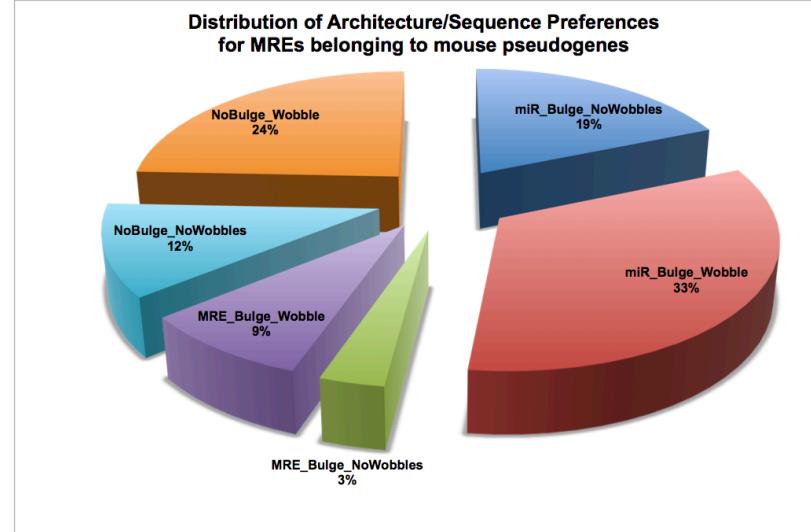
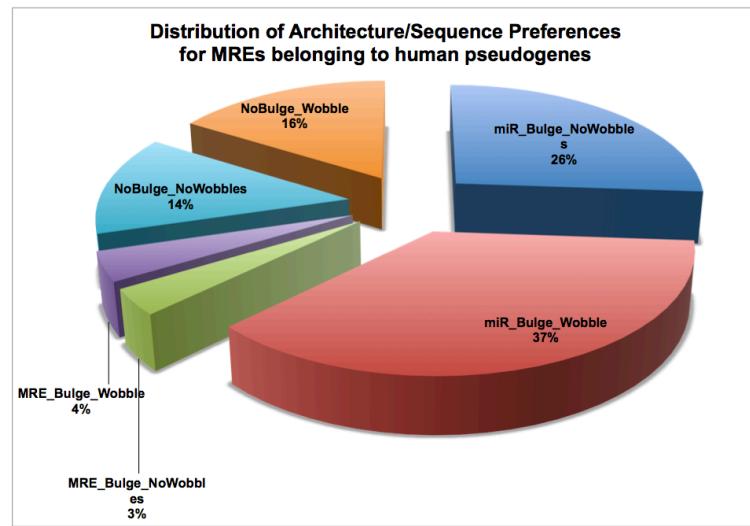
Supp. Figure 2. Example of the expected count distribution obtained from the Monte-Carlo Simulation for a single MRE motif along with the fitted negative binomial distribution.

	Human	Mouse
miR_Bulge_NoWobbles	133	697
miR_Bulge_Wobble	196	1575
MRE_Bulge_NoWobbles	22	78
MRE_Bulge_Wobble	51	326
NoBulge_NoWobbles	85	349
NoBulge_Wobble	124	1082
	611	4107



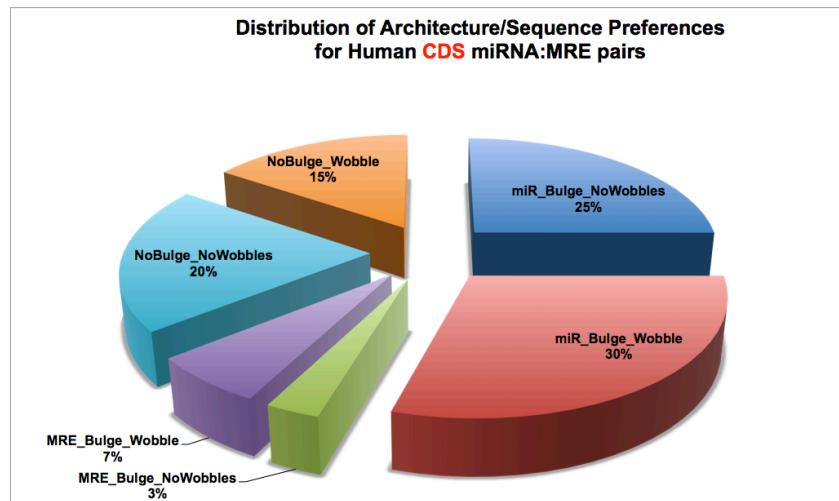
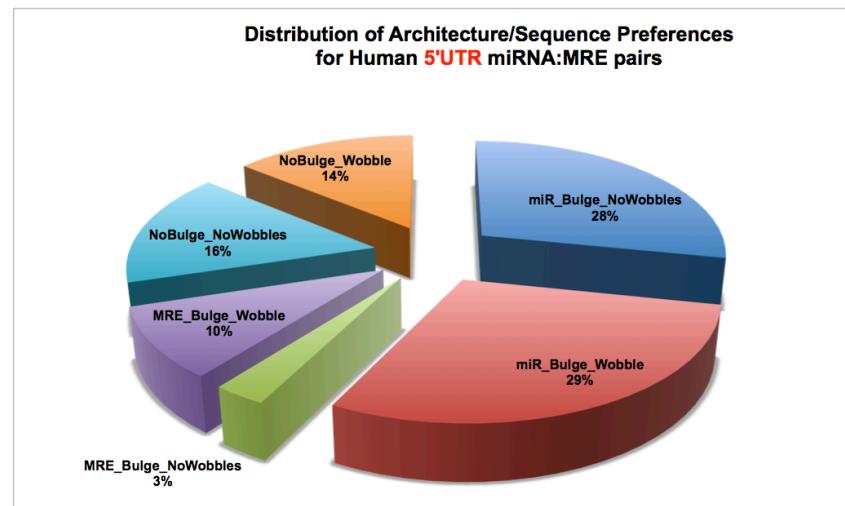
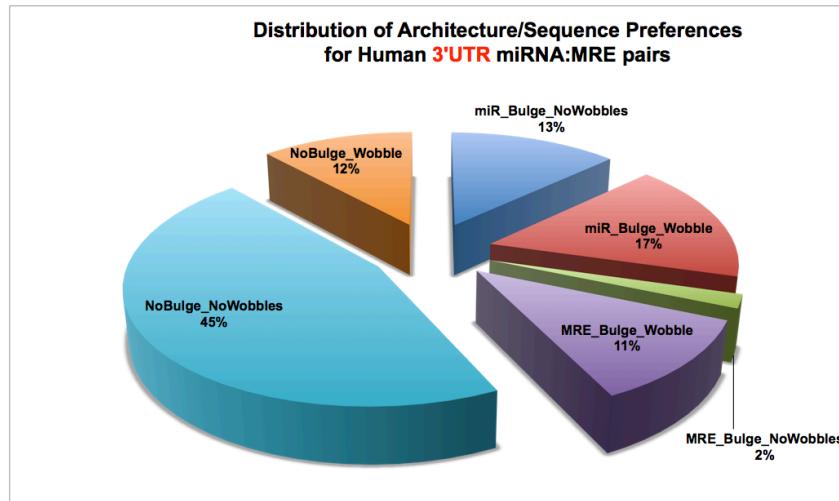
Supp. Figure 3. Number of statistically significant MREs that overlap with human and mouse IncRNAs and distribution of the corresponding elucidated architecture and sequence preferences.

	Human	Mouse
miR_Bulge_NoWobbles	52	427
miR_Bulge_Wobble	73	761
MRE_Bulge_NoWobbles	7	75
MRE_Bulge_Wobble	8	202
NoBulge_NoWobbles	27	263
NoBulge_Wobble	32	556
	199	2284



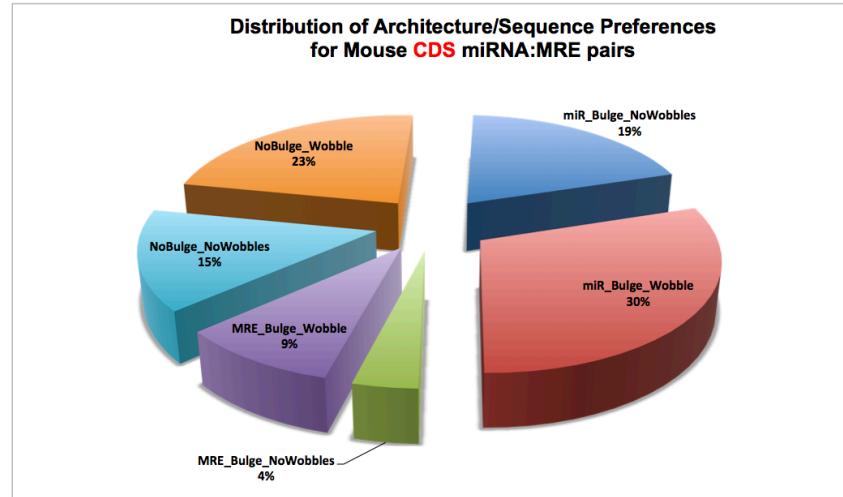
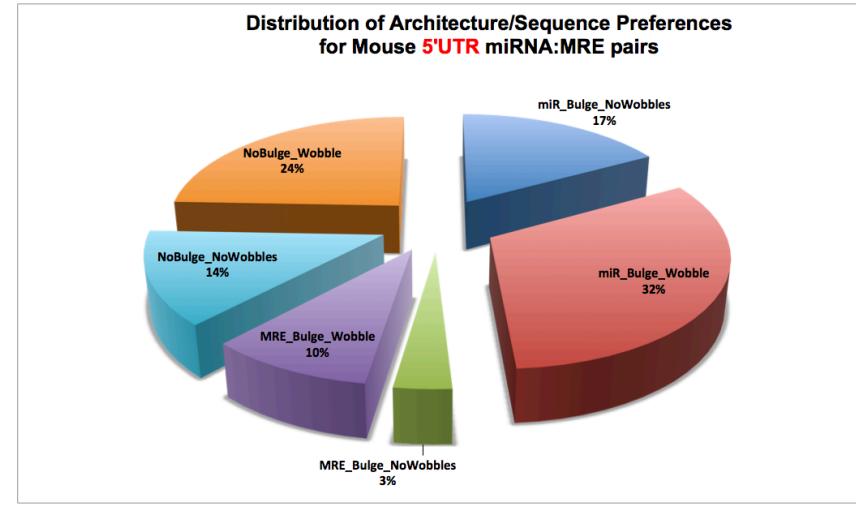
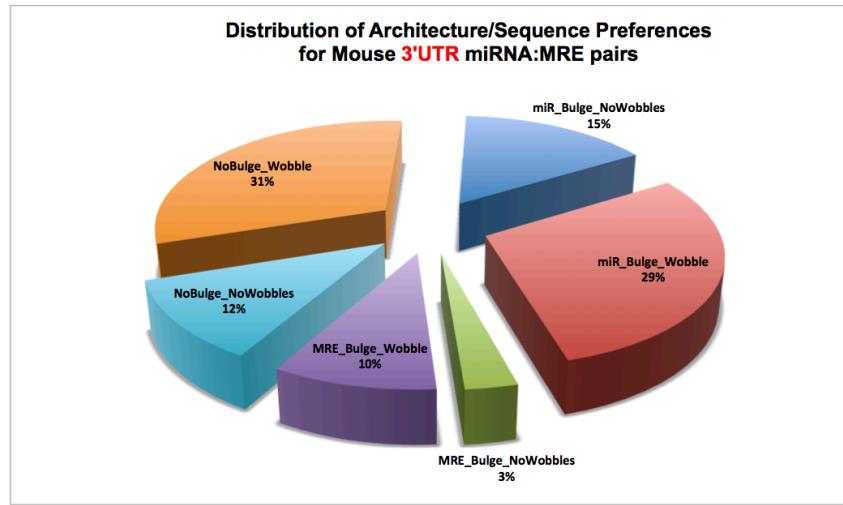
Supp. Figure 4. Number of statistically significant MREs that overlap with human and mouse pseudogenes and distribution of the corresponding elucidated architecture and sequence preferences.

	3'UTR	5'UTR	CDS
mir_Bulge_NoWobbles	900	105	1050
miR_Bulge_Wobble	1145	109	1242
MRE_Bulge_NoWobbles	130	11	126
MRE_Bulge_Wobble	793	36	277
NoBulge_NoWobbles	3147	59	853
NoBulge_Wobble	813	52	649
	6928	372	4197



Supp. Figure 5. Number of statistically significant MREs that overlap with human exons and distribution of the elucidated architecture and sequence preferences in 5'UTRs, CDSs and 3'UTRs.

	3'UTR	5'UTR	CDS
miR_Bulge_NoWobbles	43702	1414	25194
miR_Bulge_Wobble	83414	2669	39297
MRE_Bulge_NoWobbles	8668	279	4937
MRE_Bulge_Wobble	26958	802	12313
NoBulge_NoWobbles	32914	1133	19134
NoBulge_Wobble	87544	2032	29606
	283200	8329	130481

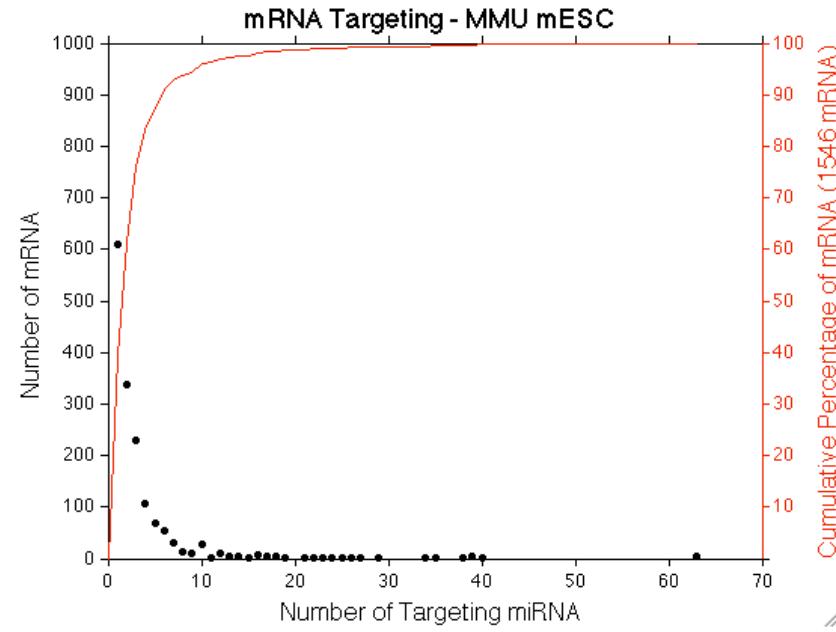
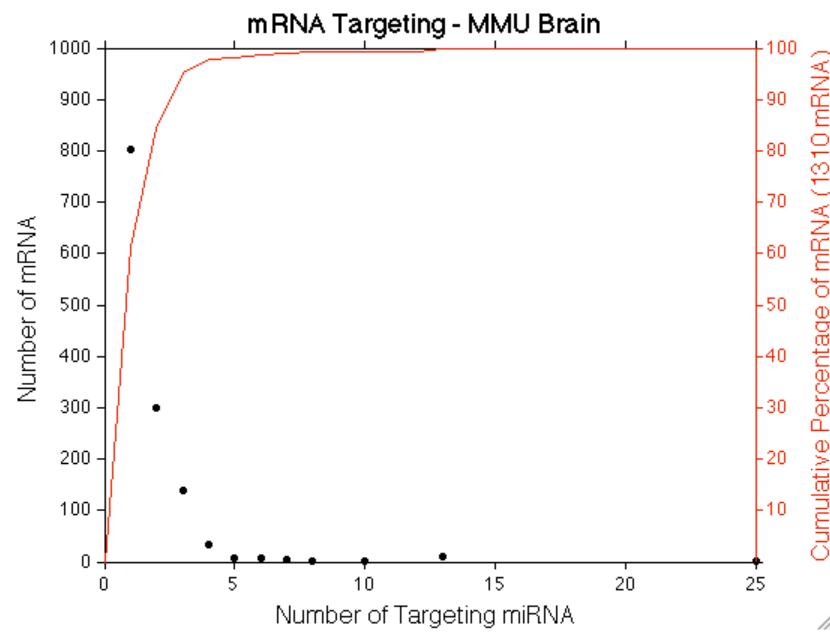


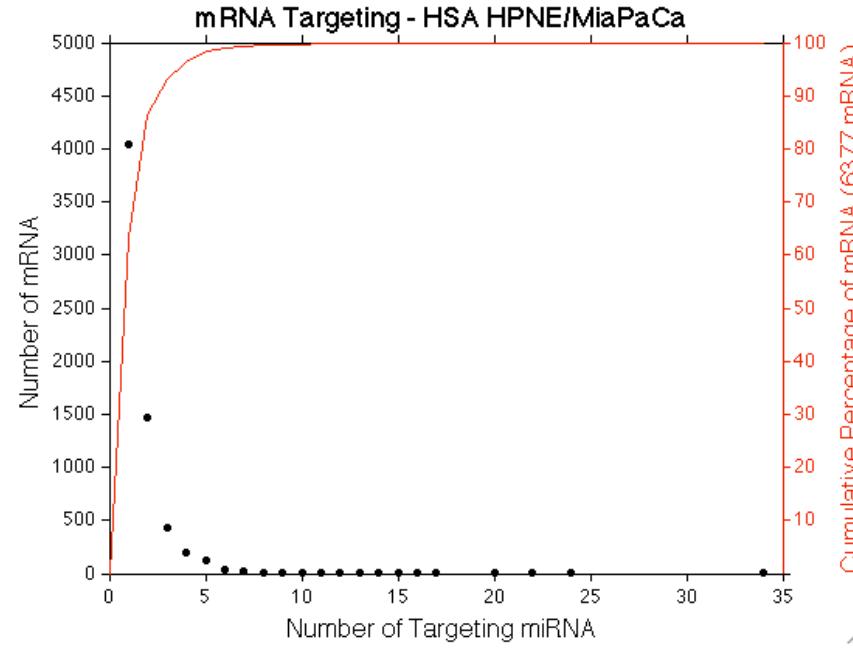
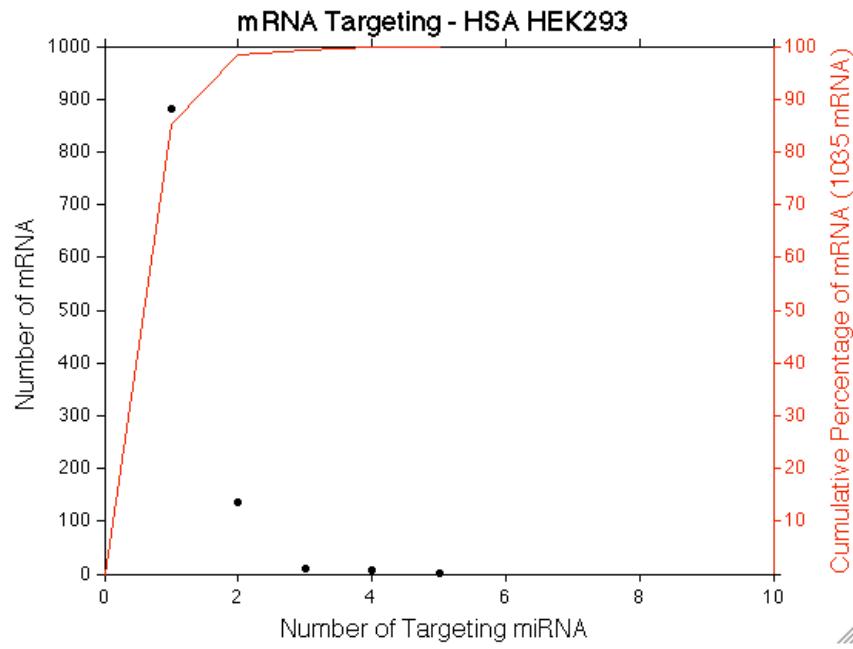
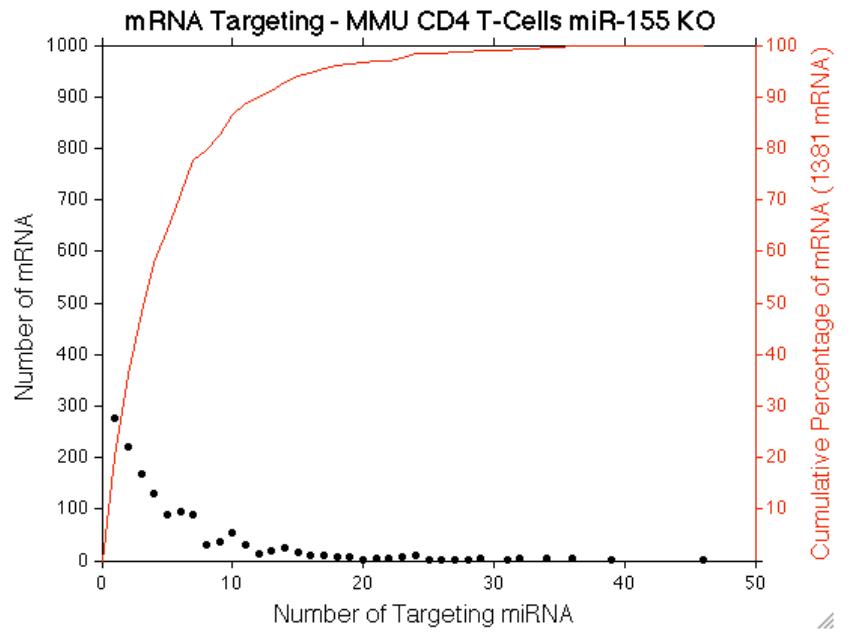
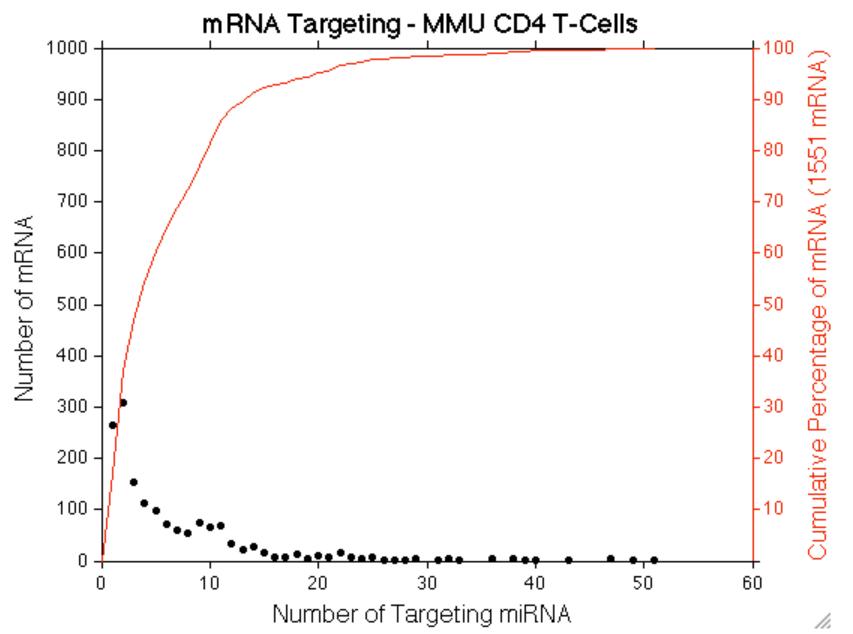
Supp. Figure 6. Number of statistically significant MREs that overlap with mouse exons and distribution of the elucidated architecture and sequence preferences in 5'UTRs, CDSs and 3'UTRs.

An mRNA can be controlled by several distinct miRNAs

Among the MREs for which we can unambiguously identify a single targeting miRNA we sub-selected those that are contained in the exons of protein-coding genes and sought to generate a *lower bound* on the number of distinct miRNAs that can target a protein-coding transcript. Our analysis reveals that the majority of distinct mRNA transcripts are targeted by one or two miRNAs: the mouse CD4⁺ T-cell samples represent an exception and their mRNAs are targeted by many more miRNAs by comparison (see below). Notably, the average number of miRNAs targeting a single mRNA transcript varied considerably across the five tissue types analyzed, which is expected considering the wide-ranging distribution of MREs across exonic and non-exonic space (Figure 3). Of the five tissue types, the highest percentage of sequenced MREs in exonic space was encountered in the mouse CD4⁺ T-cell samples. Interestingly, we find that each one of the top 10% most-regulated mRNAs are targeted by an average of 3 (mouse brain), 10 (mESC), 3 (HEK293), and 16 (mouse CD4⁺ T-cells) miRNAs, in agreement with earlier estimates.

The plots that follow show the distribution of the number of miRNAs that regulate a given mRNA transcript. Only heteroduplexes where the MREs are targeted by a single miRNA were considered for these plots. The miRNA:MRE heteroduplexes from replicates of the same tissue type have been pooled together (union of miRNA:MRE interactions across replicates from a given sample type).





REFERENCES

1. Kishore, S., et al., *A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins*. Nat Methods, 2011. **8**(7): p. 559-64.